

## BIROn - Birkbeck Institutional Research Online

Sancetta, A. and Nikandrova, Arina (2009) Forecasting and prequential validation for time varying meta-elliptical distributions. *Journal of Time Series Econometrics* 1 (2), ISSN 1941-1928.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/7036/>

*Usage Guidelines:*

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>  
contact [lib-eprints@bbk.ac.uk](mailto:lib-eprints@bbk.ac.uk).

or alternatively

# Forecasting and Prequential Validation for Time Varying Meta-Elliptical Distributions\*

Alessio Sancetta and Arina Nikandrova

Faculty of Economics, University of Cambridge, UK

April 24, 2008

## Abstract

We consider forecasting and prequential (predictive sequential) validation of meta-elliptical distributions with time varying parameters. Using the weak prequential principle of Dawid, we conduct model validation avoiding nuisance parameter problems. Results rely on the structure of meta-elliptical distributions and we allow for discontinuities in the marginals and time varying parameters. We illustrate the ideas of the paper using a large data set of 16 commodity prices.

**Keywords:** Commodity Prices, Copula Function, Meta-Elliptical Distribution, Nonparametric Estimation, Prequential Analysis, Weibull Distribution.

**JEL:**C14, C16, C31, C32.

## 1 Introduction

In this paper we describe how to model the distribution of multivariate time series using the copula. We provide a computationally intensive approach to model validation with no need to account for estimated parameters in the testing procedure. This is possible by recursive estimation of the predictive density at each point in time (or over some subsequence) using

---

\*Corresponding author: Alessio Sancetta, Faculty of Economics, Sidgwick Avenue, University of Cambridge, CB3 9DD, UK; tel. +44-1223-335364, fax. +44-1223-335399, e-mail: asancetta@gmail.com.

only preceding observations (essentially this is the real time econometrics approach considered by Pesaran and Timmermann, 2005).

We consider the following framework: each period the econometrician needs to issue a probability forecast for a vector of random variables. To issue a forecast the econometrician will select a meta-elliptical distribution parametrized by a finite dimensional vector of parameters. The dependence on the past is fully captured by the parameters that may vary over time. The proposed way to model the parameter dynamics covers parametric, semiparametric and nonparametric techniques and can be related to the existing literature (Hansen, 1994, Engle, 2002, Jondeau and Rockinger, 2005, Patton, 2006).

Having issued a forecast, the econometrician is then interested in assessing the proposed forecasting model and improving upon it. Since the forecasts are constructed recursively, the estimated parameters change over time, making the existing results on inference not directly applicable. However, by the weak prequential principle of Dawid (e.g. Dawid, 1984, 1985, 1986, Dawid and Vovk, 1999, and Seillier-Moiseiwitsch and Dawid, 1993 being very relevant), the way the forecast is constructed should be irrelevant for model validation: a model, or part of a model is retained until it is invalidated by empirical evidence. We propose a viable testing procedure that follows this principle.

There is a relatively large literature on testing for the right copula (e.g. Genest and Rivest, 1993, Breymann et al., 2003, Fermanian, 2005, Maleverge and Sornette, 2003). However, the existing studies focus on estimating a copula using sample observations and then testing for the adequacy of the postulated copula specification. Here, instead, we are in a purely forecasting framework. The model is recursively estimated and a loss is incurred every time a new observation becomes available. There are clear computational drawbacks in this approach (though it is the only one that replicates a true forecasting exercise); however, it reduces the possible problem of overfitting when estimation sample and validation sample are the same.

Using the fact that the forecasts come from a meta-elliptical distribution, we can easily transform the data so that under the null the statistics do not depend on any unknown parameters and suitable prequential validation can be carried out. The structure of the test statistics under the null is such that we only need to simulate uniform random variables, making the tabulation of critical values relatively easy. The required computations are all straightforward and relatively fast to perform in any modern machine. This approach is a

simple alternative to more analytic methods, which provide a limiting distribution for the test statistic whose critical values can be difficult to compute (e.g. Fermanian, 2005, for some solution).

The assumption of meta-ellipticity allows us to break down validation into two simple stages: marginal and copula validation. At first we check the marginal forecasts. If the marginal forecasts are not falsified, we then focus on the cross dependence structure. The advantage of breaking up validation into several stages is that this allows us to easily identify at which stage the forecasts failed. This identification issue is important for practical work.

Two-stage copula modelling is not new, but some approaches in the existing literature (either implicitly or explicitly based on the copula) lead to inconsistent distributions (e.g. Engle, 2002, Patton 2006, though the empirical performance of these methods is undisputably good). In this respect we provide some clarifications.

To illustrate the approach we consider the problem of forecasting energy and soft commodity prices. The forecasts we use are subjective, but they perform reasonably well.

Before turning to the more technical material, we briefly provide some further motivation for our forecast validation approach.

## 1.1 Further Motivation

Popper's falsifiability principle motivated this work. Loosely, it says that any theory should be potentially falsifiable. In our context, we regard the way we construct the forecast (e.g. the model and the calibration method) as the theory. Given that our forecasts are parametric, our choices of model and methods imply an underlying belief that we test on the data. For these beliefs to be falsified, it is neither necessary nor sufficient to find an alternative forecasting procedure that leads to better results. Indeed both forecasts could be falsified by reality and for this reason we need to go beyond the purely relative comparison. On the other hand if reality does not falsify either forecasting procedure, the superior relative performance of one forecasting method does not necessarily falsify the other. Indeed, one could argue that performance can be measured in many different ways depending on subjective preferences. On the other hand falsification is a concept more intimately related to the phenomenic representation of an underlying "truth", i.e. empirical observation. Hence, no forecasting method can falsify another, but the data (i.e. the phenomenical representation of the "truth") can. The objective of this paper is to take into account these remarks in

describing a procedure that allows us to evaluate forecasts in their own right. This relates to the prequential literature quoted above.

## 1.2 Plan of the Paper

The plan for the remainder of the paper is as follows. In Section 2, we introduce the class of meta-elliptic distributions and describe the important transformations of the data used to test the performance of the forecasts. In Section 3, we describe the issues involved in constructing forecasts for the marginals and the copula. In Section 4, we describe the testing procedure in details. In Section 5, we provide an illustrative example. Section 6 contains concluding remarks.

## 1.3 Notation

We introduce some notation. For typographical reasons, we may put indices in parenthesis instead of using subscripts, e.g.  $x(i)$  instead of  $x_i$ . When not needed, we may suppress subscripts. Suppose  $B$  is a set, then  $A \subset\subset B$  is a compact set inside  $B$ . For any mapping  $(x_1, \dots, x_K) \mapsto f$  and  $K > 1$ ,  $\partial_{x(k)} f := \partial f / \partial x_k$ . The inner product is denoted by  $\langle \dots, \dots \rangle$ . Furthermore, for  $X$  and  $Y$  with joint distribution  $F$ ,  $F(X|Y = y)$  or simply  $F(X|y)$  stands for the distribution of  $X$ , conditional on  $Y = y$ .

## 2 Meta-Elliptical Distributions

To define the class of meta-elliptical distributions, we first recall the definition of elliptic distributions. Let  $\mathbf{X}$  be a random vector in  $\mathbb{R}^K$ . Then  $\mathbf{X}$  is said to have an elliptic density (e.g. Kano, 1994, Fang et al., 2002) with location parameter  $\mu$ , positive definite scale matrix  $\Sigma$  and generating function  $\varphi$ , if

$$pdf_{\varphi}(\mathbf{x}) = \det |\Sigma|^{-1/2} \varphi(\langle (\mathbf{x} - \mu), \Sigma^{-1}(\mathbf{x} - \mu) \rangle). \quad (1)$$

For our purposes, we set  $\mu = \mathbf{0}$  and restrict  $\Sigma$  to have diagonal entries equal to one, i.e.  $\Sigma$  is a correlation matrix. Under this restriction on  $\Sigma$ ,  $\varphi$  is uniquely identified (e.g. Fang et al., 2002). Throughout we shall assume that  $\Sigma$  is a correlation matrix with no further mention. Under this condition, all the marginals of (1) are identical, say  $F_{\varphi}$ , and only depend on  $\varphi$

and are given by (e.g. Fang et al., 2002)

$$F_\varphi(x) = \frac{1}{2} + \frac{\pi^{(K-1)/2}}{\Gamma((K-1)/2)} \int_{-\infty}^x \int_{z^2}^{\infty} (y - z^2)^{(K-1)/2} \varphi(y) dy dz, \quad (2)$$

where  $\Gamma(x)$  is the gamma function. Fang et al. (2002) have introduced the term meta-elliptical distributions to describe distributions related to the density in (1).

For the definition of meta-elliptical distributions, we shall recall the definition of copula function. Suppose  $X_1, \dots, X_K$  are random variables with marginals  $F_1, \dots, F_K$ . Then, there is a function  $C : [0, 1]^K \rightarrow [0, 1]$  such that their joint distribution  $F$  can be written as

$$F(\mathbf{x}) = C(F_1(x_1), \dots, F_K(x_K))$$

$\mathbf{x} = (x_1, \dots, x_K)$ . When the marginals are continuous, the copula is unique and is the joint distribution of the transformed variables  $U_k = F_k(X_k)$  which are then  $[0, 1]$  uniform. We shall discuss this issue more in details in due course. Now we shall just introduce the definition of meta-elliptical copulae. We say that  $X_1, \dots, X_K$  have meta-elliptical distribution with marginals  $F_1, \dots, F_K$  and scaling function  $\varphi$  if their copula density, say  $c$ , can be written as

$$\begin{aligned} c_\varphi(\mathbf{u}; \Sigma) &= \det|\Sigma|^{-1/2} \varphi(\langle (\mathcal{Q}_\varphi \mathbf{u}), \Sigma^{-1}(\mathcal{Q}_\varphi \mathbf{u}) \rangle) J(\mathbf{u}), \\ J(\mathbf{u}) &= \left( \prod_{k=1}^K \frac{dF_\varphi^{-1}(u_k)}{du_k} \right)^{-1}, \end{aligned} \quad (3)$$

where  $\mathcal{Q}_\varphi : [0, 1]^K \rightarrow \mathbb{R}^K$  is an operator such that

$$\mathcal{Q}_\varphi \mathbf{u} = (F_\varphi^{-1}(u_1), \dots, F_\varphi^{-1}(u_K))^T. \quad (4)$$

Note that  $F_\varphi^{-1}(u) := \inf \{x \in \mathbb{R} : F_\varphi(x) \geq u\}$  is the generalized inverse of (2). Therefore, we shall call meta-elliptical copulae the class of copulae corresponding to meta-elliptical distributions. It is clear that the marginals  $F_1, \dots, F_K$  play a special role and we shall discuss the transformation  $U_k = F_k(X_k)$  and its generalization in more details within a time series context.

## 2.1 The Uniform Transform and Its Implications

To discuss the transform  $U_k = F_k(X_k)$  and its extensions in a time series context, we suppose that  $(\mathbf{X}_s)_{s \in \mathbb{N}}$  is a sequence of random vectors with values in  $\mathbb{R}^K$ . Our goal is to redefine

the random vector  $\mathbf{X}_t = (X_{t1}, \dots, X_{tK})$  into a new probability space such that each  $X_{tk}$  is transformed into a uniform  $[0, 1]$ . To achieve this, under general conditions on the Lebesgue decomposition of the marginal distributions of  $\mathbf{X}_t$ , we use the following function

$$\tilde{F}_{tk}(x, v) = \Pr(X_{tk} < x | \mathcal{F}_{t-1}) + v \Pr(X_{tk} = x | \mathcal{F}_{t-1}), \quad (5)$$

where  $\mathcal{F}_{t-1}$  is the sigma algebra generated by  $(\mathbf{X}_s)_{s < t}$ . Clearly if the conditional distribution of  $X_{tk}$  at  $x$  is continuous, it shall be equal to  $\tilde{F}_{tk}(x, v)$  no matter what  $v$  is. We state the following important result, whose proof, for example, can be found in Rio (2000, Lemma 1F).

**Lemma 1** *Suppose  $V_{tk}$  is a  $[0, 1]$  uniform random variable. Then,*

$$U_{tk} := \tilde{F}_{tk}(X_{tk}, V_{tk})$$

*is a  $[0, 1]$  uniform random variable independent of  $\mathcal{F}_{t-1}$  and*

$$X_{tk} = \inf \{x \in \mathbb{R} : \Pr(X_{tk} \leq x | \mathcal{F}_{t-1}) \geq U_{tk}\}$$

*almost surely.*

Define

$$\mathbf{U}_t = (U_{t1}, \dots, U_{tK}) = \left( \tilde{F}_{t1}(X_{t1}, V_{t1}), \dots, \tilde{F}_{tK}(X_{tK}, V_{tK}) \right), \quad (6)$$

where  $(V_{tk})_{t \in \mathbb{N}}$  is iid mutually independent for each  $k$  and independent of  $(\mathbf{X}_t)_{t \in \mathbb{N}}$ . By Lemma 1,  $(\mathbf{U}_t)_{t \in \mathbb{N}}$  is a sequence of iid random variables with uniform  $[0, 1]$  marginals, i.e.  $\mathbf{U}_t$  is independent of  $\mathcal{F}_{t-1}$ . Since  $\mathbf{U}_t$  is a random vector with uniform marginals, its joint distribution is the copula of  $\mathbf{X}_t$  conditional on  $\mathcal{F}_{t-1}$ . (For convenience, in the sequel we shall suppress the restriction on  $\mathcal{F}_{t-1}$  both in the formulae as well as in the text). If  $\mathbf{X}_t$  does not have marginals that are continuous, there is not a unique copula for it (e.g. Sklar, 1973 Corollary to Theorem 1). However,  $\mathbf{U}_t$  is derived from (6) and it is a continuous random variable. Hence, there is a unique copula for  $\mathbf{U}_t$  which is also its joint distribution. If the copula of  $\mathbf{X}_t$  is not unique, the copula of  $\mathbf{U}_t$  will be used as a unique choice of copula. This solves any identification issues for the copula of  $\mathbf{X}_t$ , and the term copula would only refer to this version with continuous uniform marginals. We now turn on some implications of the fact that  $\mathbf{U}_t$  is independent of  $\mathcal{F}_{t-1}$ .

### 2.1.1 A Pitfall in Copula Modelling

The fact that, by Lemma 1,  $\mathbf{U}_t$  is independent of  $\mathcal{F}_{t-1}$  is often ignored in applied work leading to inconsistent models based on the copula. We consider a very simple illustrative example.

**Example 2** *As usual  $\mathbf{X}_t = (X_{t1}, \dots, X_{tK})$  and  $\mathcal{F}_{t-1}$  is the sigma algebra generated by its past. Let  $X_{tk} = Z_{tk}\sigma_{tk}$  be such that  $(Z_{tk})_{t \in \mathbb{Z}}$  is a sequence of iid standard Gaussian random variables and  $(\sigma_{tk})_{t \in \mathbb{Z}}$  is a sequence such that  $\sigma_{tk}$  is  $\mathcal{F}_{t-1}$  measurable. Note that*

$$\int_{-\infty}^x \frac{1}{\sigma} \phi\left(\frac{s}{\sigma}\right) ds = \int_{-\infty}^{x/\sigma} \phi(s) ds$$

where  $\phi$  is the standard Gaussian density. Then, it is easy to see that

$$F_{X_{tk}}(X_{tk}|\mathcal{F}_{t-1}) = F_{Z_{tk}}(Z_{tk}|\mathcal{F}_{t-1})$$

where  $F_{X_{tk}}(\cdot|\mathcal{F}_{t-1})$  and  $F_{Z_{tk}}(\cdot|\mathcal{F}_{t-1})$  are the conditional distributions of  $X_{tk}$  and  $Z_{tk}$  respectively. Given that  $(Z_{tk})_{t \in \mathbb{Z}}$  is iid, then,  $Z_{tk}$  is independent of  $\mathcal{F}_{t-1}$ . Now suppose that  $\mathbf{X}_t$  conditioning on  $\mathcal{F}_{t-1}$  has Gaussian copula. By the properties of Gaussian copulae, the scaling matrix is equal to the covariance of  $\mathbf{Z}_t := (Z_{t1}, \dots, Z_{tK})$  conditional on  $\mathcal{F}_{t-1}$ . But  $\mathbf{Z}_t$  is independent of  $\mathcal{F}_{t-1}$  (it is iid) hence, the conditional covariance (correlation) equals the unconditional covariance. Moreover, GARCH modelling of the marginals together with a copula to model the cross sectional dependence often leads to some more obvious inconsistency. In fact, one first models the marginals conditioning on  $(X_{sk})_{s < t}$  and then joins them together assuming a copula conditioning on  $\mathcal{F}_{t-1}$ . There is nothing wrong with this, but it is clear that the parameters of the copula cannot depend on  $\mathcal{F}_{t-1}$  otherwise also the marginals would depend on this larger sigma algebra (Fermanian and Wegkamp, 2004, for details and interpretations).

The above example shows that Engle (2002) DCC model is inconsistent. Indeed, there the correlation matrix depends on  $\mathcal{F}_{t-1}$ . Using the independence of  $\mathbf{U}_t$  and  $\mathcal{F}_{t-1}$  it follows that other models based on the copula are inconsistent. For example Patton (2006) is right in his analysis suggesting that one should use the conditional copula. He is wrong in modelling the parameters of the conditional copula as  $\mathcal{F}_{t-1}$  dependent. Despite this theoretical inconsistency these models have been very successful in modelling the joint distribution of assets' returns. A possible reason is that when estimated, these models are able to capture time



inhomogeneity of the copula parameters (i.e. nonstationarity). For example the exponential smoothing used by Engle (2002) to estimate the correlation matrix of the DCC model can act as a filter tracking a time inhomogeneous correlation matrix that may change slowly in time. (Time inhomogeneous distributions for assets' returns have been advocated by Stărică and Granger, 2005).

We shall conclude this section with a further transformation that pertains only to meta-elliptical distributions and that together with the uniform transform will be used for testing purposes.

## 2.2 Scaling Matrix Transforms

One important property of  $\mathbf{U}_t$  having meta-elliptical copulae with full rank scaling matrix  $\Sigma_t$  is that we can write  $\mathbf{Z}_t := \mathcal{Q}_\varphi \mathbf{U}_t$ , where  $\mathbf{Z}_t$  has elliptical distribution with scaling matrix  $\Sigma_t = A_t A_t^T$  for some  $K \times K$  matrix  $A_t$  (the decomposition  $\Sigma_t = A_t A_t^T$  is unique only up to a multiplicative orthonormal matrix). This implies the following stochastic representation (e.g. Hult and Lindskog, 2002),

$$\mathbf{Z}_t := \mathcal{Q}_\varphi \mathbf{U}_t = R \mathbf{S} \quad (7)$$

where the equality is in distribution and  $A$  is a  $K \times K$  matrix such that  $AA^T = \Sigma$ ,  $R$  is a positive random variable uniquely determined by  $\varphi$  (because  $\Sigma_t$  is a correlation matrix) and  $\mathbf{S}$  is  $K$  dimensional uniformly distributed on the unit sphere and independent of  $R$ .

**Example 3** *Let  $\mathbf{Z}_t$  be  $K$  dimensional standard Gaussian. Then  $\mathbf{Z}_t = R\mathbf{S}$  where  $R^2$  is Chi square with  $K$  degrees of freedom (e.g. Hult and Lindskog, 2002, Example 3.1).*

Representation (7) implies two simple but important properties of meta-elliptical copulae which we state next, using the notation in (7) with no further mention. The proof is elementary and left out.

**Lemma 4** *Suppose  $\mathbf{U}_t$  has meta-elliptical copula with generator  $\varphi$  and scaling matrix  $\Sigma$ . Then,*

$$R = |A^{-1}\mathbf{Z}_t|$$

and

$$\mathbf{S} = \frac{A^{-1}\mathbf{Z}_t}{|A^{-1}\mathbf{Z}_t|}.$$

Moreover,

$$\mathbf{U}'_t := \mathcal{Q}_\varphi^{-1} A^{-1} \mathbf{Z}_t \quad (8)$$

has same meta-elliptical copula as  $\mathbf{U}_t$  but with scaling matrix  $I_K$ , the  $K$  dimensional identity matrix.

From Lemma 4 and (7) it is easy to see that the random variable  $R$  (which determines  $\varphi$ ) affects the dependence properties of meta-elliptical copulae. The goal of the paper is to find ways to transform random variables so that they do not depend on any specific parameter. Parameters' invariance is important because the distribution of our test statistics will be approximated by simulation; this will be come clear in due course. If we try to orthogonalize the random variables using the scaling matrix  $\Sigma$  we obtain a distribution that is independent of unknown parameters. In the next examples, orthogonalization is equivalent to assuming that  $\Sigma = I_K$ .

**Example 5** Define  $\varphi(x) := \varphi^K(x) = \exp(-\frac{x}{2}) / (2\pi)^{K/2}$ , which is the generator for the Gaussian copula. Then,

$$c_\varphi(\mathbf{u}; I_K) = \frac{1}{(2\pi)^{K/2}} \exp\left(-\frac{\langle(\mathcal{Q}_\varphi \mathbf{u}), (\mathcal{Q}_\varphi \mathbf{u})\rangle}{2}\right) J(\mathbf{u}),$$

where

$$J(\mathbf{u}) = (2\pi)^{K/2} \exp\left(\frac{\langle(\mathcal{Q}_\varphi \mathbf{u}), (\mathcal{Q}_\varphi \mathbf{u})\rangle}{2}\right).$$

Hence,  $c_\varphi(\mathbf{u}; I_K) = 1$ , which is the independence copula and the orthogonalization leads to a copula that does not depend on any unknown parameters.

Unfortunately, this is not always the case.

**Example 6** Define

$$\varphi(x) := \varphi_\gamma^K(x) = \frac{\Gamma((\gamma + K)/2)}{\Gamma(\gamma/2) (\pi\gamma)^{K/2}} (1 + x/\gamma)^{-(\gamma+K)/2},$$

which is the generator for the  $t$ -copula with  $\gamma$  degrees of freedom. Then,

$$c_\varphi(\mathbf{u}; I) = \frac{\Gamma((\gamma + K)/2)}{\Gamma(\gamma/2) (\pi\gamma)^{K/2}} \left(1 + \frac{\langle(\mathcal{Q}_\varphi \mathbf{u}), (\mathcal{Q}_\varphi \mathbf{u})\rangle}{\gamma}\right)^{-(\gamma+K)/2} J(\mathbf{u})$$

$$J(\mathbf{u}) = \prod_{k=1}^K \frac{\Gamma(\gamma/2) (\pi\gamma)^{K/2}}{\Gamma((\gamma + 1)/2)} \left(1 + \frac{[F_\varphi^{-1}(u_k)]^2}{\gamma}\right)^{(\gamma+1)/2},$$

and it is simple to see that  $c_\varphi(\mathbf{u}; I) \neq 1$  because it depends on  $\gamma$ .

By Lemma 4 we can transform the variables so that they have distribution  $R$ . We can then use Lemma 1 to obtain iid uniform random variables. We shall now turn to the forecasting procedure so that it will be clear why in that framework the parameters used in the forecasts clearly change over time.

### 3 Methodology

The focus of this paper is not in identifying the "true" model, but in model validation in terms of recursive forecast performance.

Suppose  $(\mathbf{X}_s)_{s \in \mathbb{N}}$  is a sequence of random vectors with values in  $\mathbb{R}^K$ . We observe realizations from the segment  $(\mathbf{X}_s)_{0 \leq s < t}$ , i.e.  $\mathbf{x}_0, \dots, \mathbf{x}_{t-1}$ . We need to issue a probability forecast for  $\mathbf{X}_t$ . In the absence of knowledge about the (conditional) distribution of  $\mathbf{X}_t$ , we could postulate a family of joint distributions indexed in some finite parameter space and choose a method to identify the optimal parameters for the forecasting problem. For large  $K$ , this is a formidable task, unless we are prepared to impose very restrictive conditions on the family of distributions and the dynamics of the parameters. We propose to consider the problem of providing a good forecast for each marginal distribution first and then to join these marginals using some nonlinear transformations and the copula function.

It is very important to note that we consider admissible forecasting: at each point in time, the forecast to be used at time  $t$  is constructed only using past observations up to time  $t - 1$ . This means that the parameters used in the forecast cannot be estimated looking at the whole sample size. By construction, this makes the estimated parameters time varying, either because of estimation error or because they might vary over time in addition to any possible estimation error. We provide details of this approach and then suggest a way to conduct inference in this case.

#### 3.1 First Stage: The Marginals

Postulate a class of marginal distributions  $\{P_\theta, \theta \in \Theta \subset \mathbb{R}^d, d \geq 1\}$  to be used for probability forecasts of  $(X_{tk})_{t \in \mathbb{N}}$ . Then, we need to choose an  $\mathcal{F}_{t-1}$  measurable  $\theta_t$  such that  $P_{\theta(t)}$  is a good distribution forecast for  $X_{tk}$ . It is customary to let  $X_{tk}$  depend on  $\mathcal{F}_{t-1}$  only through its own past values, i.e.  $(X_{sk})_{0 \leq s < t}$  rather than  $(\mathbf{X}_s)_{0 \leq s < t}$ . This will be often assumed with no need of further mention. For the moment, to simplify notation in the discussion, let

$d = 1$ . Suppose there is some function  $g : \mathbb{R} \rightarrow \mathbb{R}$  such that  $\theta_t = \mathbb{E}(g(X_t) | \mathcal{F}_{t-1})$ . Many distributions/processes satisfy this condition, e.g. GARCH ( $g(x) = x^2$ ) and more generally the exponential family model (van Garderen, 1997).

Suppose  $g(X_t)$  admits the semimartingale representation

$$g(X_t) = f_t + v_t \varepsilon_t, \quad (9)$$

where  $f_t$  and  $v_t$  are  $\mathcal{F}_{t-1}$  measurable and  $\mathbb{E}\varepsilon_t = 0$ ,  $\mathbb{E}\varepsilon_t^2 = 1$ . Hence,  $\theta_t = f_t$  (by reparametrization of the marginal distributions, this covers the case  $g'(\theta_t) = f_t$ , for some function  $g'$ ). Then the estimation of  $\theta_t$  is equivalent to estimation of the  $\mathcal{F}_{t-1}$  measurable trend in  $g(X_t)$ . We can estimate or at least approximate  $\theta_t$  by  $\hat{\theta}_t = \sum_{s < t} w(s, t) g(x_s)$ , where  $(w(s, t))_{0 \leq s < t}$  is a linear filter possibly depending on  $(X_s)_{s < t}$  so that  $(w(s, t))_{0 \leq s < t}$  is  $\mathcal{F}_{t-1}$  measurable. This framework encompasses many different methods like averages, moving averages, exponential smoothing (Harvey, 1989, Muth, 1960), kernel smoothing and linear projections (i.e. OLS).

**Example 7** Suppose  $w(s, t) = (1 - h)h^{|t-s-1|}$ , which is just exponential smoothing, and  $g(x) = x^2$ . This is RiskMetrics volatility estimation.

By choice of the linear filter, we can account for a time heterogenous parameter if we suspect that this is a problem.

**Example 8** Suppose  $X_t = \sigma_t Z_t$  where  $Z_t$  is standard Gaussian, and  $\sigma_t$  is  $\mathcal{F}_{t-1}$  measurable. For volatility estimation, Mercurio and Spokoiny (2004) propose for  $\alpha > 0$

$$\mathbb{E}(|X_t|^\alpha | \mathcal{F}_{t-1}) = (\#\mathcal{I})^{-1} \sum_{s \in \mathcal{I}} |X_s|^\alpha$$

over a homogeneity region  $\mathcal{I}$  (arguing for  $\alpha = 1/2$  as the best choice). The linear filter gives same weight over the homogeneity region and zero elsewhere. Mercurio and Spokoiny (2004) provide inference procedures to identify homogeneity regions.

More generally than (9), we can suppose that there is a function  $g : \mathbb{R} \times \Theta \rightarrow \mathbb{R}$  such that  $\mathbb{E}(g(X_t, \theta_t) | \mathcal{F}_{t-1}) = 0$ , and  $g(X_t, \theta)$  admits the semimartingale representation

$$g(X_t, \theta) = f_t(\theta) + v_t(\theta) \varepsilon_t,$$

where  $\mathbb{E}(g(X_t, \theta) | \mathcal{F}_{t-1}) = f_t(\theta)$ . Applying a filter  $(w(s, t))_{0 \leq s < t \in \mathbb{N}_+}$  such that  $\hat{f}_t(\theta) = \sum_{s < t} w(s, t) g(X_s, \theta)$ , numerical approximation techniques can be employed to find  $\hat{\theta}_t$ , where  $\hat{\theta}_t$  is such that  $\hat{f}_t(\hat{\theta}_t) = 0$ .

The  $d > 1$  dimensional case is dealt similarly either by defining a vector of estimating equations or by direct solution if the parameter defining equation admits an explicit solution as a function of  $X_t$ .

As the last step we join the marginals and derive the full joint distribution. To conduct inference, we shall make use of (7) hence attention is restricted to the class of meta-elliptical copulae. Next we discuss parameters' choice in a way that is consistent with Lemma 1.

### 3.2 Second Stage: The Copula

Restrict attention to elliptical copulae with given function  $\varphi$ . We suppose that  $\mathbf{U}_t$  has copula given by  $c_\varphi(\mathbf{u}; \Sigma_t)$  where, by the previous remarks about  $\mathbf{U}_t$ ,  $\Sigma_t$  is independent of  $\mathcal{F}_{t-1}$ . Hence  $\Sigma_t$  is either a deterministic function of  $t$ , i.e. measurable with respect to  $\mathcal{F}_{-\infty} = \bigcap_{s=1}^{\infty} \mathcal{F}_{t-s}$ , or depends on  $\mathbf{U}_t$  only. As mentioned in Section 2.1.1, this fact is often overlooked in the applied literature, where the scaling matrix  $\Sigma_t$  is understood to be a conditional correlation or in general a parameter depending on  $\mathcal{F}_{t-1}$ .

Despite this remark, models like Engle's DCC appear to be quite successful suggesting that, while  $\Sigma_t$  is independent of  $\mathcal{F}_{t-1}$ , it might be the case that  $\Sigma_t$  changes slowly with  $t$  (i.e. it is time inhomogeneous). In this case, we may still try to track its values via some calibration method. We assume that there exists a function  $g : \mathbb{R}^K \rightarrow \mathbb{R}^K \times \mathbb{R}^K$  such that  $\Sigma_t = \mathbb{E}g(\mathbf{U}_t)$ , as  $\mathbf{U}_t$  is independent of  $\mathcal{F}_{t-1}$ . (We do not need to condition on  $(V_{tk})_{t \in \mathbb{N}}$  used in (6) because this is an iid sequence independent of  $(\mathbf{X}_t)_{t \in \mathbb{N}}$  for each  $k$ ). The use of meta-elliptical copulae implies that  $g(\mathbf{u}) = (\mathcal{Q}_\varphi \mathbf{u})(\mathcal{Q}_\varphi \mathbf{u})^T$ . In many cases,  $\varphi = \varphi_\gamma$  where  $\gamma$  is some parameter in a compact parameter space, and this may complicate matters.

**Example 9** *For the  $t$ -copula, the generator  $\varphi$ , as given in Example 6, depends on the degrees of freedom  $\gamma$ , i.e.  $\mathcal{Q}_\varphi$  depends on  $\gamma$ .*

In these cases, we may assume a specific value for  $\gamma$ . Alternatively, the scaling matrix can still be obtained from Kendall's tau, say  $\rho_\tau$ . Kendall's tau is a measure of dependence like the correlation coefficient, but it is independent of the marginal distributions. For the pair of random variables  $(U_i, U_j)$

$$\rho_\tau := 4\mathbb{E}I_{\{U_j \leq U_i\}}I_{\{V_j \leq V_i\}} - 1 = \mathbb{E} \text{sign}(U_i - U_j) \text{sign}(U'_i - U'_j),$$

where  $(U'_i, U'_j)$  is an independent copy of  $(U_i, U_j)$  (Joe, 1997, for more details). It is well known (Lindskog et al., 2003, Fang et al., 2002) that for a meta-elliptical copula

$$\rho = \sin\left(\frac{\pi\rho_\tau}{2}\right), \quad (10)$$

where  $\rho$  is the usual correlation coefficient. Hence, for arbitrary meta-elliptical copulae we can always suppose that  $\arcsin(\Sigma_t) = \mathbb{E}g(\mathbf{U}_t)$ , where the  $(i, j)$  entry of  $g$  is  $\pi\rho_{\tau;i,j,t}/2$ , i.e. proportional to Kendall's tau for  $U_{ti}$  and  $U_{tj}$ . To approximate the expectation, we use a linear filter, so that, mutatis mutandis, this framework is similar to the marginals modelling. A nonparametric estimator for  $\rho_{\tau;i,j,t}$  is given by a second order U-process

$$\hat{\rho}_{\tau;i,j,t} = \sum_{0 \leq r < s < t} w(r, s, t) \operatorname{sign}(U_{ir} - U_{jr}) \operatorname{sign}(U_{is} - U_{js}), \quad (11)$$

where the filter  $(w(r, s, t))_{0 \leq r < s < t}$  is deterministic.

**Example 10** Suppose  $w(r, s, t) = \binom{t}{2}^{-1}$ . This is the simplest case, where we assume homogeneity and (11) reduces to an ordinary U-statistic.

**Example 11** Suppose

$$w(r, s, t) = B_t^{-1} k_h(t-r) k_h(t-s), \quad B_t := \sum_{0 \leq r < s < t} k_h(t-r) k_h(t-s),$$

where  $k_h(s) := h^{-1}k(h^{-1}s)$  and  $k(s)$  is a decreasing positive function. Hence, the most recent observations are given the largest weight in an attempt to track slow variation in  $\Sigma_t$ . For  $k_h(s) = (1-h)h^{|s|-1}$  with  $h \in (0, 1)$  this becomes a double exponential smoothing.

Since  $(\mathbf{U}_t)_{t \in \mathbb{N}}$  is a sequence of iid random vectors, the results in Goshal et al. (2000) can be used to study the properties of local versions of Kendall's tau. When  $\varphi$  is fully known, we may avoid working with a second order U-process and resume more direct methods, i.e.

$$\hat{\Sigma}_t = \sum_{s \leq t-1} w(s, t) (\mathcal{Q}_\varphi \mathbf{U}_s) (\mathcal{Q}_\varphi \mathbf{U}_t)^T,$$

Whatever the method employed, any additional parameter can be calibrated subsequently once  $\hat{\Sigma}_t$  has been estimated by direct method or Kendall's tau.

Finally, we note that by (7) and the remarks in Section 2.2,  $R$  uniquely determines  $\varphi_\gamma$ , which we assume to be indexed by the finite dimensional parameter  $\gamma$  (see Sancetta, 2007, for the general nonparametric problem). Then there is a direct link between  $\gamma$  and the distribution of  $R$  and mutatis mutandis, the arguments in Section 2.2 apply.

**Example 12** Suppose  $\mathbf{U}_t$  has a  $K$  dimensional  $t$  copula with  $\gamma$  degrees of freedom and scaling matrix  $I_K$ . Then,  $R_t^2 = \gamma\chi_{(K)}^2/\chi_{(\gamma)}^2$  in distribution, where  $\chi_{(s)}^2$  is a Chi-square random variable with  $s$  degrees of freedom. Then, a forecast of  $\gamma$  is standard. For example, when  $\gamma > 2$ , we can use (9) with

$$\mathbb{E}g(R_t^2/K) = \mathbb{E}R_t^2/K = \frac{\gamma}{\gamma - 2},$$

because  $R_t^2/K$  is  $F$  distributed with  $(K, \gamma)$  degrees of freedom, hence its mean is  $\gamma/(\gamma - 2)$ . Time inhomogeneity can be considered as well.

### 3.3 Optimization of the Linear Filter

The linear filters used in the forecasts usually depend on some parameters that may need to be calibrated.

**Example 13** Suppose  $(w(s, t))_{0 \leq s < t} = (w(s, t))_{h \leq s < t} = (\sum_{h \leq r < t} Z_{r,h} Z_{r,h}^T)^{-1} \sum_{0 \leq r < t} Z_{r,h} X_r$  where  $Z_{r,h} = (X_{r-1}, \dots, X_{r-h})$  (In this case, it is more convenient to give the whole filter). Then, the filter is just a linear projection on the  $h$  past values, the autoregressive order.

In the above example, the filter depends on some unknown parameter  $h$ . To produce valid forecasts, the filtering parameters to be used at time  $t$  might be chosen to minimize the forecast error

$$\mathcal{E}_{t-1}(h) := \sum_{1 \leq s \leq t-1} \mathcal{R}(f_s - \hat{f}_{s,h}),$$

where  $\hat{f}_{s,h}$  is the estimated trend at time  $s - 1$  (where the filter depends on some finite parameter vector  $h$ ),  $f_s$  is the true trend and  $\mathcal{R}(x)$  is a convex function of  $x$ . Clearly,  $f_s$  is unobservable, and has to be replaced by  $g(X_s)$  suggesting minimization of the estimated forecast error

$$\mathcal{E}'_{t-1}(h) := \sum_{1 \leq s \leq t-1} \mathcal{R}(g(X_s) - \hat{f}_{s,h}). \quad (12)$$

Define  $\tilde{h}_t = \arg \min \mathcal{E}_{t-1}(h)$  and  $\hat{h}_t = \arg \min \mathcal{E}'_{t-1}(h)$ . Cheng et al. (2003) show that with probability going to one, as  $t \rightarrow \infty$ ,  $\mathcal{E}_{t-1}(\hat{h})$  performs as well as  $\mathcal{E}_{t-1}(\tilde{h})$  when  $\mathcal{R}(x) = x^2$ , or  $|x|$ , under regularity conditions. Hence the parameters in the filter can be estimated as the ideal choice of filtering parameters. Other approaches include forecast combination via learning algorithms (e.g. Yang, 2004, Sancetta 2006, 2007). We shall use this latter approach in the illustrative empirical example.

## 4 Prequential Validation

Once a forecasting procedure has been chosen (this comprises of the parametric model and the calibration strategy for the parameters), we shall see if it can be rejected by the data. To this end, we suppose that we have a sample  $X_{1k}, \dots, X_{nk}$ ,  $k = 1, \dots, K$  and their associated marginals probability forecasts. (here, it is more convenient to use a different indexing for the random variables, i.e. not  $0, \dots, t-1$ , but  $1, \dots, n$ ). Using our forecasts for the probabilities and (6) we obtain  $\hat{U}_{1k}, \dots, \hat{U}_{nk}$ ,  $k = 1, \dots, K$ . The hat stresses the fact that we are using the forecasted probabilities in (6). To avoid trivialities in the indexing, we assume that at time 0 a probability forecast is already available. It is worth stressing once again that the forecasts only depend on the past, so that the estimation is recursive and all parameters used at iteration  $i$  shall use previous observations only.

Marginals validation involves testing if  $\hat{U}_{1k}, \dots, \hat{U}_{nk}$  is iid  $[0, 1]$  uniform for each  $k$  (by Lemma 1). This implies that the data are generated by the same marginal conditional probabilities that we are using to issue the forecast. Since true conditional probabilities are unknown and forecasts calibrated through learning are used to approximate them, it is unlikely that  $\hat{U}_{1k}, \dots, \hat{U}_{nk}$  is uniform. However, as argued in Seillier-Moiseiwitsch and Dawid (1993) we may abstract from the way the forecasts are generated and evaluate them on the basis of their performance only: this is the basis of the weak prequential principle (e.g. Dawid, 1997). If the forecasts were not such that  $\hat{U}_{1k}, \dots, \hat{U}_{nk}$  is (approximately) iid uniform, this hypothesis would be rejected by the data and our forecasting procedure discredited (i.e. falsified using the language of the introduction).

If this hypothesis is not discredited, then we can go to the next level and check that  $\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_n$  has elliptic copula, where  $\hat{\mathbf{U}}_i := (\hat{U}_{i1}, \dots, \hat{U}_{iK})$  and its copula parameters are denoted by  $\hat{\Sigma}_i = \hat{A}_i \hat{A}_i^T$  and  $\hat{\gamma}_i$  (assuming  $\varphi = \varphi_\gamma$  fully determined by  $\gamma$ , e.g. Example 12). To ease notation we let  $\hat{\varphi}_i := \varphi_{\hat{\gamma}_i}$ . Following Lemma 4, define

$$\hat{R}_i := \left| \hat{A}_i^{-1} \left( \mathcal{Q}_{\hat{\varphi}_i} \hat{\mathbf{U}}_i \right) \right|, \hat{\mathbf{S}}_i := \frac{\hat{A}_i^{-1} \left( \mathcal{Q}_{\hat{\varphi}_i} \hat{\mathbf{U}}_i \right)}{\left| \hat{A}_i^{-1} \left( \mathcal{Q}_{\hat{\varphi}_i} \hat{\mathbf{U}}_i \right) \right|} \quad (13)$$

and let  $F_{\hat{\gamma}_i}$  be the distribution of  $\hat{R}_i$  implied by  $\hat{\gamma}_i$ . For the sake of explanation assume  $F_{\hat{\gamma}_i}$  is continuous. Then the copula forecast should not be discredited if  $\hat{\mathbf{S}}_1, \dots, \hat{\mathbf{S}}_n$  is uniformly distributed on the sphere and  $\hat{U}_1^{(R)}, \dots, \hat{U}_n^{(R)}$ , where  $\hat{U}_i^{(R)} := F_{\hat{\gamma}_i} \left( \hat{R}_i \right)$ , is a sequence of iid  $[0, 1]$  uniform. The procedure is identical to the marginal case.



**Example 14** *In Example 12, let  $F_\gamma$  be the  $F$  distribution with  $(K, \gamma)$  degrees of freedom. Then,  $F_{\hat{\gamma}_i}(\hat{R}_i/K)$  should be close in distribution to a uniform random variable for the forecast not to be discredited.*

It follows that for "good" forecasts, the null distribution of the random variables does not depend on any unknown parameter. In fact all random variables are transformed to  $[0, 1]$  uniform random variables or to  $K$  dimensional uniforms on the unit sphere. Without these transformations it would be difficult to conduct any inference as the estimated quantities depend on unknown parameters (under the null) that are recalibrated at each step, as required by any admissible forecasting procedure.

Note that it can be challenging to test for a  $K$  dimensional random vector to be uniformly distributed on the unit sphere. However, let  $W_1, \dots, W_n$  be iid Chi square random variables with  $K$  degrees of freedom. Following Example 3,  $W_1\hat{\mathbf{S}}_1, \dots, W_n\hat{\mathbf{S}}_n$  should be approximately standard Gaussian random vectors if the forecasts are "good".

Before giving some details, we mention that when the number of hypotheses is large, usual critical values should not be used. Indeed if we consider testing for  $M$  hypotheses we shall have uniform control of all of them and not just one of them at the time. This is the usual data snooping problem: as  $M$  becomes large, we will always have at least one rejection. There are procedures to control for multiple hypotheses and in this framework, the ones based on Bonferroni inequality and Holm's step down procedure appears to be more easily applicable (see Romano et al., 2008, for details).

## 4.1 Some Details

In this section, we provide details of the testing procedures implemented in the empirical example. All test statistics are described for illustrative purposes and are selected based on subjective preference. In general we will define some empirical process  $\xi_n(\mathbf{u})$  where in the univariate case  $\mathbf{u}$  is just a scalar. This will be a properly standardized version of the empirical distribution function of the transformed variables (details are given below, but for the sake of unity we describe the general form of the test statistic now). For statistical applications,

interest lies in the  $L_p$  norm of the empirical process  $\xi_n(\mathbf{u})$  i.e.

$$\left| \int_{[0,1]^K} |\xi_n(\mathbf{u})|^p d\mathbf{u} \right|^{1/p}, p \in [1, \infty) \quad (14)$$

$$\sup_{\mathbf{u} \in (0,1)^K} |\xi_n(\mathbf{u})|, p = \infty.$$

While for the cases we consider the asymptotic distribution of  $\xi_n(\mathbf{u})$  is standard, the distribution of the  $L_p$  norm is not easy to derive and thus we suggest a simulation approach. Furthermore, it is not easy to compute  $L_p$  norm of  $\xi_n(\mathbf{u})$ . However, the  $L_p$  norm can be found by Monte Carlo integration : choose  $M$  large and generate  $(\mathbf{U}_i)_{i \in \{1, \dots, M\}}$  uniform random variables and compute

$$\left[ \frac{1}{M} \sum_{i=1}^M |\xi_n(\mathbf{U}_i)|^p \right]^{1/p} \quad (15)$$

with the obvious modifications when  $p = \infty$ . The transformations described in this paper allow us to easily simulate the test statistics. Since we explicitly test for a series of null hypotheses, we describe each of them separately.

Test the null  $H_0^{(1)}$ :  $(\hat{U}_{ik})_{i \in \mathbb{N}}$  is a sequence of iid uniform  $[0, 1]$  random variables against the alternative  $H_1^{(1)}$ :  $(\hat{U}_{ik})_{i \in \mathbb{N}}$  is not  $[0, 1]$  uniform. Define:

$$\xi_n^{(1)}(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( I\{\hat{U}_{ik} \leq u\} - u \right)$$

and construct its  $L_p$  norm by Monte Carlo methods as in (15). Given that under the null  $(\hat{U}_{ik})_{i \in \mathbb{N}}$  is iid uniform, we can compute critical values simulating samples of iid uniform random variables. For each sample we construct  $\xi_n^{(1)}(u)$ , compute its  $L_p$  norm (as in (15)), and find the critical values from the empirical quantiles. While computationally intensive, the procedure does not require challenging asymptotics, as it relies on simulating uniform random variables.

The test for  $H_0^{(1)}$  might have low power. For example, not only it is necessary that  $(\hat{U}_{ik})_{i \in \mathbb{N}}$  is a sequence of uniform random variables, but we also require that they are independent. We may still fail to reject  $H_0^{(1)}$  in the presence of some quite weak dependence. Hence, it is worthwhile to test the null hypothesis  $H_0^{(2)}$ : the sequence of uniform random variables  $(\hat{U}_{ik})_{i \in \mathbb{N}}$  is independent against the alternative  $H_1^{(2)}$ : the uniform random variables are not independent. We are assuming that the random variables are uniform, which is plausible if the null  $H_0^{(1)}$  in the previous paragraph is not rejected.

There is a plethora of tests for independence (see Diebold et al., 1998, and Patton, 2006 for interesting examples). It is quite common to carry out inference for finite number of lags (and in the sequel only lag one is considered for the sake of simplicity) using the sample autocovariance function of the series and/or some simple functions of the series (e.g. powers of the centered series). This does not necessarily implies independence unless we consider  $Cov\left(\hat{U}_{ik}^p, \hat{U}_{i+1,k}^q\right)$  for all  $p, q \in \mathbb{N}$ . Since the variables are in  $[0, 1]$ ,  $Cov\left(\hat{U}_{ik}^p, \hat{U}_{i+1,k}^q\right) \rightarrow 0$  as  $p, q \rightarrow \infty$  (unless there is some positive mass at 1), looking at a finite number of combinations of  $p$  and  $q$  might be sufficient.

An alternative that fits within the empirical distribution framework is to look at the empirical distribution of  $\left(\hat{\mathbf{U}}_i^{(l)}\right)_{i>l} = \left(\hat{U}_i, \hat{U}_{i-1}, \dots, \hat{U}_{i-l}\right)_{i>l}$ . To test time series independence at the first  $l$  lags, we set  $C_I(\mathbf{u}) = \prod_{i=1}^l u_i$ , and construct the following statistic under the null of independence:

$$\xi_n^{(2)}(\mathbf{u}) := \sum_{i=1}^n \left( I\left\{ \hat{\mathbf{U}}_i^{(l)} \leq \mathbf{u} \right\} - C_I(\mathbf{u}) \right). \quad (16)$$

Again the distribution of the variables is fully specified under the null and we can use simulations to compute critical values for the  $L_p$  norm of  $\xi_n^{(2)}(\mathbf{u})$ . Since  $l$  is finite, it is clear that the test does not have power against variables that are dependent at high lags, but independent at low lags. However, that would be an odd dependence structure.

If we cannot discredit the marginals forecasts, we then turn to the copula. Recall the notation in (13) and how we find these variables. Then, test the null  $H_0^{(3)} : \left(\hat{U}_i^{(R)}\right)_{i \in \mathbb{N}}$  is a sequence of iid uniform  $[0, 1]$  random variables against the alternative  $H_1^{(3)} : \left(\hat{U}_i^{(R)}\right)_{i \in \mathbb{N}}$  is not  $[0, 1]$  uniform as well as the null  $H_0^{(4)} : \left(\hat{\mathbf{S}}_i\right)_{i \in \mathbb{N}}$  is a sequence of iid uniform random variables on the  $K$  dimensional sphere against the alternative  $H_1^{(4)} : \left(\hat{\mathbf{S}}_i\right)_{i \in \mathbb{N}}$  is not uniformly distributed on the sphere. Simulating a sequence  $(W_i)_{i \in \mathbb{N}}$  of iid Chi square random variables with  $K$  degrees of freedom we can amend the null distribution to standard Gaussian for the sequence  $\left(W_i \hat{\mathbf{S}}_i\right)_{i \in \mathbb{N}}$ . This shows that in some special simple cases, the procedure can be simplified considerably.

**Example 15** Suppose  $\mathbf{U}_t$  has Gaussian copula with scaling matrix  $\Sigma_t = A_t A_t^T$ . Then, the generator  $\varphi$  does not depend on any unknown parameter and

$$\mathbf{Z}_t := A_t^{-1} \mathcal{Q}_\varphi^{-1} \mathbf{U}_t \quad (17)$$

is a standard normal vector. Hence we can combine  $H_0^{(3)}$  and  $H_0^{(4)}$  into a unique hypothesis for Gaussianity of  $\left(\hat{\mathbf{Z}}_i\right)_{i \in \mathbb{N}}$  where

$$\hat{\mathbf{Z}}_t = \hat{A}_i^{-1} \left( \mathcal{Q}_\varphi^{-1} \hat{\mathbf{U}}_i \right)$$

and  $\varphi$  is known.

Note that  $\xi_n(\mathbf{u})$  does not have constant variance. We could standardize by its standard deviation, say  $\sigma(\mathbf{u}, \mathbf{u})$ , but this leads to instability in the tails, because  $\sigma(\mathbf{u}, \mathbf{u}) \rightarrow 0$  as  $\mathbf{u} \rightarrow \partial[0, 1]^K$ , where  $\partial[0, 1]^K$  is the boundary of  $[0, 1]^K$ . For this reason, we discuss the unstandardized process  $\xi_n(\mathbf{u})$ , as done for the Kolmogorov-Smirnov statistic. Clearly, this implies less powerful tests.

## 5 Empirical Illustration with Soft Commodities and Energy Prices

We use an empirical example to illustrate the procedure in a practical situation. In particular we focus on forecasting commodity prices. There has been some interest in applied work on commodity prices (Taylor, 1980, Deaton and Laroque, 1992, 1996, Deb et al., 1996, Pindyck, 2004).

The forecasting approach might look ad hoc. However, we take care to check if our preferred (very subjective) method to construct forecasts is discredited by the data. This has no implication for any alternative procedure.

### 5.1 Data Description and in Sample Analysis

We use futures data on 16 different commodities over the period 3/01/93-28/01/04, which amounts to about 2700 observations excluding festivities. The data were retrieved from Bloomberg using the generic ticker for each contract including festivities. This delivers a concatenated series of front months for each contract series, where the price is the previous period price in the case of a closed market. For each commodity, the front month has a life of about one to three months. We eliminated all festivities and computed log return for all series as  $R_t := \ln(P_t/P_{t-1})$ . Every time the front month changed at time  $t$ ,  $P_t$  and  $P_{t-1}$  referred to two different contracts (in terms of delivery). Unfortunately, contracts on

different commodities postulate different delivery periods (for example soybeans and coffee have different settlement dates) so that it is difficult to avoid this problem for a large data set. However, visual inspection of our data did not reveal any clear short term seasonal pattern that might arise from the roll over of contracts.

The commodities studied are crude oil, gas oil (IPE), heating oil, natural gas, propane, unleaded gas, cocoa, coffee, sugar, orange juice, soybean, corn, rice, oats, wheat and cotton. Assuming the data possess suitable ergodic properties, we report sample summary statistics in Table I.

Table I.

Table I. Sample Summary Statistics

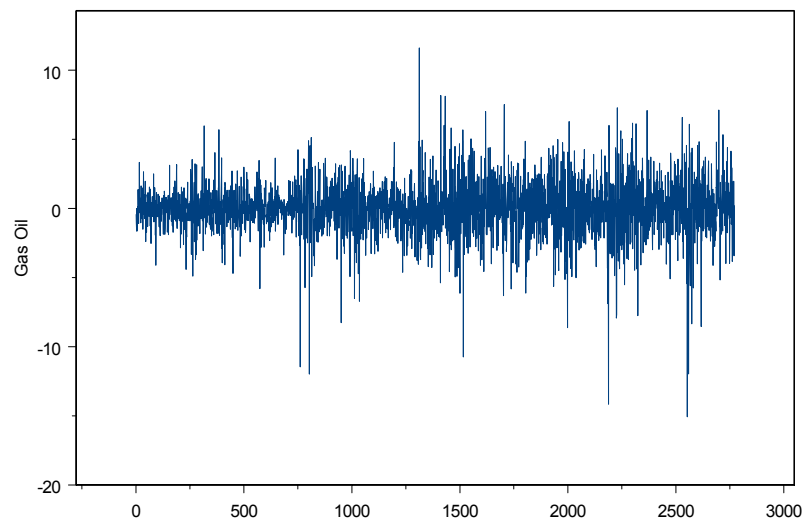
	Mean	Var.	Skew.	Kurt.	Min.	1stQu.	Median	3rdQu.	Max.
CRUDE OIL	0.02	5.18	-0.27	4.07	-16.54	-1.15	0.00	1.29	14.23
GAS OIL (IPE)	0.01	3.96	-0.55	5.17	-15.07	-0.98	0.00	1.06	11.62
HEATING OIL	0.02	5.38	-0.95	7.35	-20.97	-1.13	0.03	1.29	10.40
NATURAL GAS	0.04	14.53	-0.21	7.84	-37.57	-1.95	0.05	2.03	32.44
PROPANE	0.03	4.14	-1.26	15.84	-24.78	-0.80	0.00	0.92	12.18
UNLEADED GAS	0.02	5.77	-0.59	7.74	-25.45	-1.27	0.10	1.41	19.49
COCOA	0.02	3.82	0.25	2.72	-10.01	-1.11	0.00	1.08	12.74
COFFEE	0.00	8.06	0.09	7.04	-22.06	-1.40	0.00	1.40	23.77
SUGAR	-0.01	4.76	-0.18	10.76	-18.04	-1.05	0.00	1.07	23.55
ORANGE JUICE	-0.01	4.25	0.72	11.28	-12.91	-0.99	0.00	0.94	22.72
SOYBEANS	0.01	1.86	-1.02	13.28	-17.43	-0.73	0.03	0.76	7.41
CORN	0.01	2.23	-1.84	36.15	-26.12	-0.74	0.00	0.76	9.13
RICE	0.01	3.56	1.03	23.15	-12.97	-0.89	0.00	0.89	28.08
OATS	0.00	4.89	-1.91	20.18	-25.46	-0.92	0.00	1.02	14.54
WHEAT	0.00	3.20	-1.39	41.86	-28.61	-1.00	0.00	0.97	23.30
COTTON	0.01	3.06	-1.45	36.48	-30.44	-0.88	0.00	0.86	13.62

The results are consistent with what is expected from fat-tailed non-Gaussian data sets. Figure I gives the time series plot for the log returns of gas oil, coffee and rice.

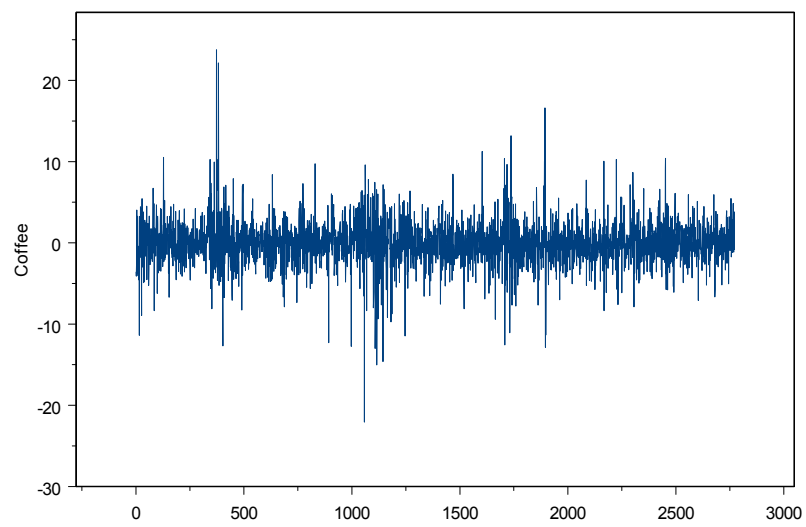
Figure I

Figure I. Time Series Plot.

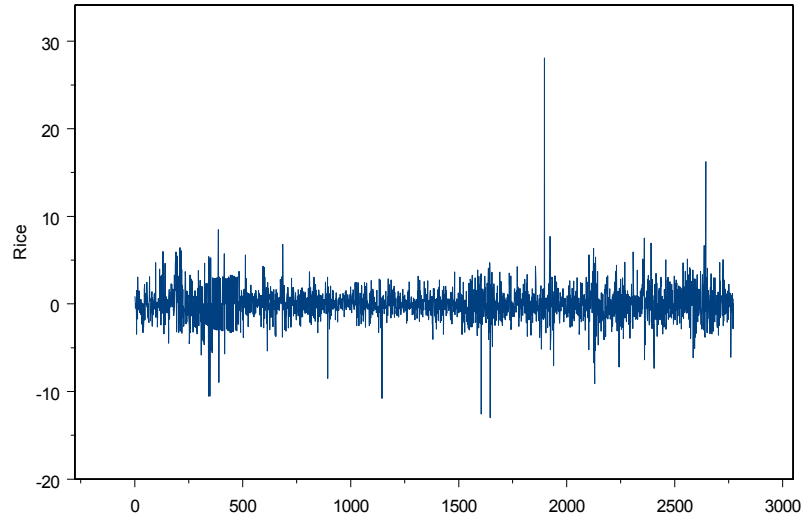
Gas Oil



Coffee



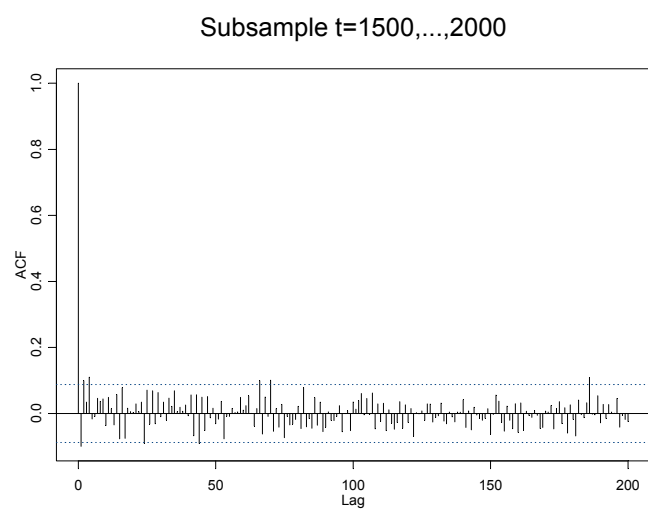
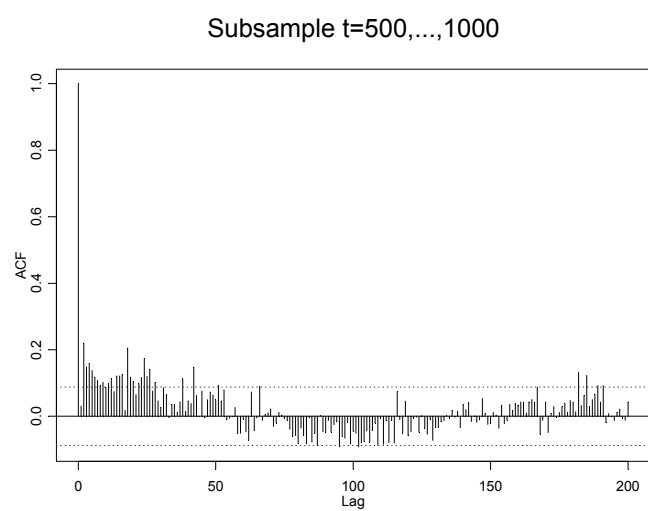
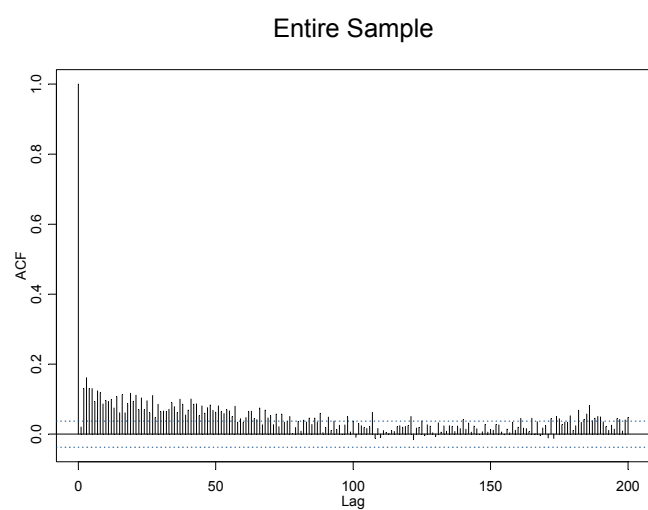
## Rice



We also looked at the sample autocorrelation function (ACF) of the data, which appeared to be consistent with conditionally heteroskedastic models (e.g. GARCH): no correlation in the levels, but correlation in absolute powers of the returns, as for financial returns. To further investigate the time dependence structure we considered estimation of subsamples of the sample ACF in search of possible anomalies. Figure II show that the sample ACF for absolute returns of gas oil seem to depend on the sample subset used, revealing possible nonstationarities in the data or the non-existence of a second moment. The same anomalies are also shared by most of the other commodities.

Figure II

Figure II. Sample ACF for Gas Oil





For the purposes of cross-sectional analysis, commodities were divided in three groups: energy, miscellaneous and grains/oilseeds/fiber. For soft commodities (i.e. non-energy group), this division is based on the nomenclature used by the Economic Research Service tariff database of the United States Department of Agriculture. The energy group comprises of crude oil, gas oil, heating oil, natural gas, propane, unleaded gas; miscellaneous group consists of cocoa, coffee, sugar and orange juice; grains/oilseeds/fiber include soybeans, corn, rice, oats, wheat and cotton. Our subdivision of soft commodities is justified by the fact that correlation of two commodities coming from different groups is usually very low.

The overall correlation among different energy commodities is rather strong. However, energy commodities exhibit virtually no correlation with commodities of other groups. Correlation between different commodities within miscellaneous group is rather weak (.3-.11), and there is virtually no correlation between miscellaneous commodities and the last group. The correlation among grains/oilseeds/fiber is lower than correlation among energy commodities, but higher than the correlation within the miscellaneous commodities. For economy of space these results will not be reported here. However, we will report the covariance matrices once we transform the data using the forecast estimator of the transform in (6). This will allow us to compare correlations within groups before and after the application of the second transform (8) used to test for meta-Gaussianity. Therefore, we now introduce the model we use to forecast the marginals.

## 5.2 A Simple Model

We propose a simple model for the forecast of the distribution of the commodities returns. As done in Stărică and Granger (2005), we decide to model the absolute returns and their sign separately. Suppose  $(R_t)_{t \in \mathbb{Z}}$  is a sequence of random variables with values in  $\mathbb{R}$ . This sequence is interpreted as log returns. We use the following decomposition  $R_t = \varepsilon_t |R_t|$ ,  $\varepsilon_t = \text{sign}(R_t)$ . In several occasions, our series has zero returns on some days. Therefore,  $\Pr(\varepsilon_t = 0) > 0$ , and the distribution function of  $R_t$  has a discrete component at zero in its Lebesgue decomposition. The distribution of  $\varepsilon_t$  is modelled as  $\Pr(\varepsilon_t = 0) = p$ ,  $\Pr(\varepsilon_t > 0) = q_t$ , and  $q_t$  can change over time. This allows for trends in prices. Trends for commodities might be the effect of seasonal patterns and other exogenous variables like weather. Moreover, for simplicity we do not model directly the break that occurs every time there is a roll of a contract. (As mentioned above this happens every one or three months depending on the

contract specification.)

We suppose that there is an  $\mathcal{F}_{t-1}$  measurable function  $\theta_t$  such that  $(|R_t|/\theta_t|\mathcal{F}_{t-1})$  (i.e.  $|R_t|/\theta_t$  conditional on the past) has Weibull density with shape parameter  $\delta$ , i.e.  $\delta x^{\delta-1} \exp\{-x^\delta\}$ . This functional form can approximate the tails of a large number of density functions and is sometimes used in the literature (e.g. Silvapulle and Granger, 2001, Laherrère and Sornette, 1998, Frisch and Sornette, 1997). The conditional mean of returns is usually very small at all frequencies but the lower ones. Note that under this assumption,  $\sigma_t := \mathbb{E}_{t-1}|R_t| = \theta_t \Gamma(1 + 1/\delta)$  and we can reparametrize in terms of  $\sigma_t$  (the conditional mean of  $|R_t|$ ) to find

$$pdf_{|R_t||\mathcal{F}_{t-1}}(x) = [\sigma_t \Gamma(1 + 1/\delta)]^\delta \delta x^{\delta-1} \exp\left\{-[\sigma_t \Gamma(1 + 1/\delta) x]^\delta\right\}.$$

To join the marginal distributions, we choose a Gaussian copula where the scale matrix is allowed to vary over time. This is done for the sake of simplicity. Inferential procedures with the Gaussian copula do not depend on nuisance parameters (i.e. the function  $\varphi$  is fully determined). Hence it is appropriate to use it as a benchmark model, and to look for better alternatives if the null of properly specified copula is rejected. Moreover, this copula is simple to estimate. Given lack of more detailed information, but only the scaling matrix, it is known that the Gaussian distribution is optimal in an information theoretic sense (i.e. it is maximum entropy).

### 5.2.1 Forecasting

In order to forecast, we consider recursive estimation of the parameters and use the observations up to time  $t - 1$  to forecast at time  $t$ . In particular parameters are either set to some a priori value or exponential moving averages are used to approximate conditional means. The smoothing parameter in the moving average is also chosen a priori. The quality of the forecast is assessed using the inferential procedures discussed in the previous section.

Note that a convenient feature of the model is that  $q_t$  and  $p$  can be estimated separately from the other parameters. We use the empirical mean of  $I\{R_s = 0\}$  (using  $s < t$ ) to estimate  $p$  and exponential moving average with smoothing parameter  $h$  to estimate  $q_t$ , i.e.  $q_t = (1 - h)q_{t-1} + hI\{R_{t-1} > 0\}$ . In the sequel we will use  $h$  to generically denote the smoothing parameter in exponential moving average estimation of any time varying parameters. A priori we chose  $h = .98$  for  $q_t$  and this is somewhat arbitrary. We conjectured

that there is little predictability in the sign of returns, so to avoid capturing noise, a large value of smoothing parameter was chosen. To estimate  $\sigma_t$  we consider exponential moving average with  $h = .95, .85$  resulting in two different estimates which we will then combine using combination weights. Two choices are also considered for the shape parameter:  $\delta = 1, 1.25$ . The first corresponds to the exponential density (e.g. Silvapulle and Granger, 2001) the latter is motivated by our past experience of working with financial returns, suggesting that choices between 1 and 1.5 are plausible but with  $\delta$  closer to one. Instead of choosing one model among these 4 possible combinations (i.e. two choices for  $\sigma_t$  and two for  $\delta$ ), we consider a forecast combination of densities via the linear opinion pool (Genest and Zidek, 1986, Timmermann, 2006):

$$pdf_{|R_t||\mathcal{F}_{t-1}}(x) = \sum_{i=1}^4 w_{ti} pdf_{|R_t||\mathcal{F}_{t-1}}(x|i),$$

where each  $i$  identifies a density for  $|R_t||\mathcal{F}_{t-1}$  with a given  $h$  for  $\sigma_t$  and a fixed value for  $\delta$ . The weights  $w_{ti}$  are constrained to lie in the unit simplex. Details on how the weights are computed can be found in the Appendix A.

Because of the Meta-Gaussianity assumption, we only need to estimate the scaling matrix. Using (6) and the estimated marginals at time  $t$ , we derive  $\hat{\mathbf{U}}_t$  and  $\hat{\mathbf{Z}}_t := \mathcal{Q}_\varphi \hat{\mathbf{U}}_t$  where  $\mathcal{Q}_\varphi$  is as in (4) and  $\varphi$  is as in Example 5. Then, the scaling matrix  $\hat{\Sigma}_h(t)$  for the Gaussian copula is estimated by exponential moving average of  $\left(\hat{\mathbf{Z}}_s \hat{\mathbf{Z}}_s^T\right)_{s < t}$  with two choices of  $h = .95, .85$ . Hence, we assume that the scaling matrix changes slowly with time. This estimator may take values outside  $[-1, 1]$  and be singular. To avoid these problems we use the following crude approach. Define  $\hat{\sigma}_{kl}$  to be the  $(k, l)$  entry in  $\hat{\Sigma}_h(t)$ . Replace  $\hat{\sigma}_{kl}$  with  $\hat{\sigma}_{kl} / (\hat{\sigma}_{kk} \hat{\sigma}_{ll})^{1/2}$ , so that the diagonal entries of the estimator are all one and the off diagonal entries are between one and minus one. Then shrink the estimator towards the identity matrix using shrinkage parameter equal to .1, which reflects our confidence in the exponential moving average estimator versus a constant diagonal one. We then use combination weights to average over the two estimators  $\hat{\Sigma}_h(t)$  (based on two different smoothing parameters). Appendix A gives details on how these second set of combination weights were chosen.

While there are many steps involved, the estimation is fully recursive. This minimizes the computational burden, particularly in the case of large number of assets in online estimation. Clearly, if this procedure were to be applied to a very large number of assets, the scaling matrix would have to be homogeneous and estimated using a small number of predetermined

factors.

Our goal is to conduct prequential validation using this forecasting model. To this end the transforms (6) and (8) from time  $t = 1, \dots, n$  use the forecasting model suggested based on previous observations only. Hence, we fully replicate a forecasting exercise. Clearly, we already know the data and we have analyzed them in the previous subsection. Nevertheless, the simplicity of the forecasting model tends to avoid any strong supposition of data mining in this forecasting and inferential analysis.

### 5.3 Analysis of the Marginals

We transformed data into uniform using (6) and the estimated marginals. Then, the Kolmogorov-Smirnov (K-S) statistic was computed for all 16 commodities. Results are in Table II from which we infer that the model forecasts all the data well, but for heating oil, propane and rice. To avoid any possible accusation of data mining, we decided not to amend the forecasts at this stage. Hence we dropped rice from our subsequent analysis, while the other two commodities were retained as we failed to reject  $H_0^{(1)}$  at the 99% confidence level.

Table II.

Table II. Test for Goodness of Fit

	K-S Stat
CRUDE OIL	0.019
GAS OIL (IPE)	0.023
HEATING OIL	0.031 **
NATURAL GAS	0.021
PROPANE	0.032 **
UNLEADED GAS	0.021
COCOA	0.021
COFFEE	0.018
SUGAR	0.019
ORANGE JUICE	0.018
SOYBEANS	0.022
CORN	0.021
RICE	0.047 ***
OATS	0.018
WHEAT	0.026
COTTON	0.021

\* rejected at 90%

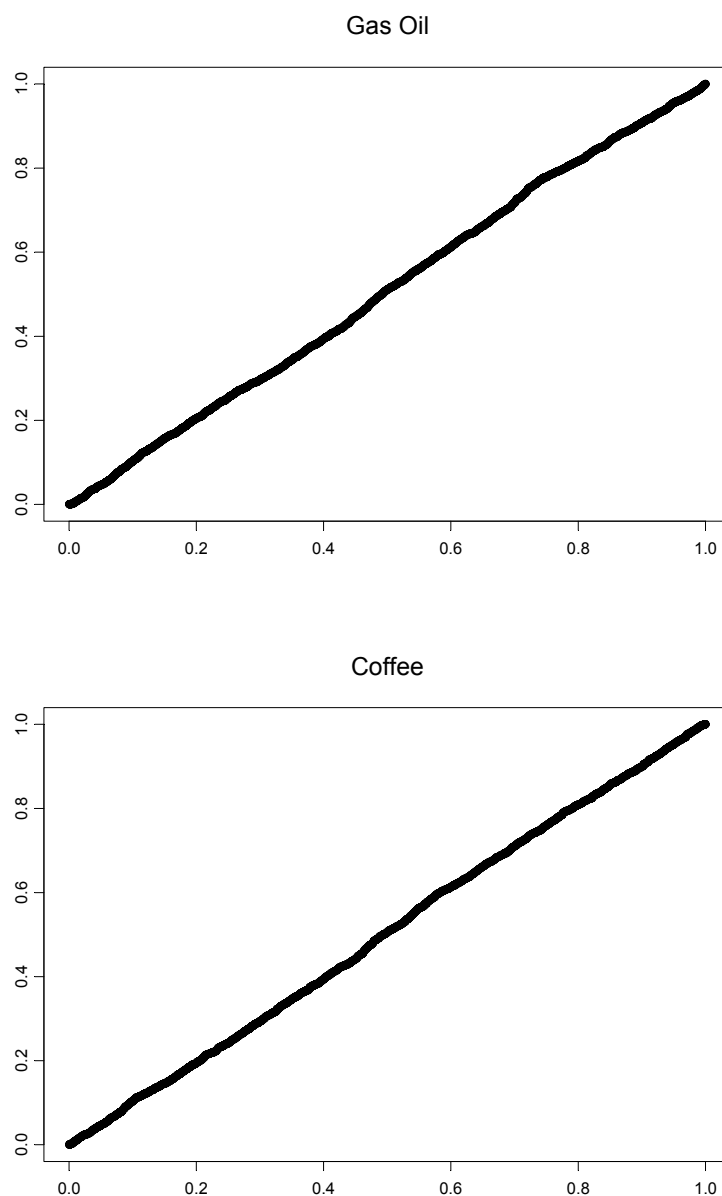
\*\* rejected at 95%

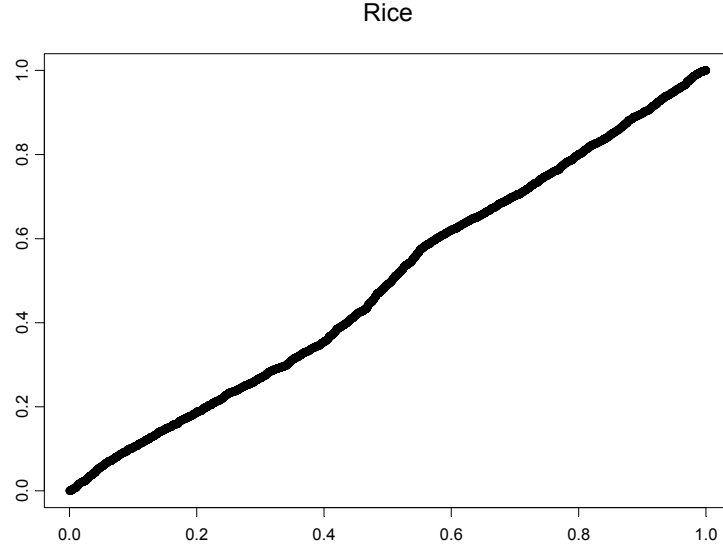
\*\*\* rejected at 99%

To provide visual evidence, Figure III reports the QQ-plots for the uniform transform of gas oil, coffee and rice (the worse performing commodity according to our statistics). The graphs show a good fit for the first two and mediocre for the latter. However, for rice the problem appears to be in the body of the distribution rather than on the tails.

Figure III.

Figure III. QQ-Plot Estimated Distribution.



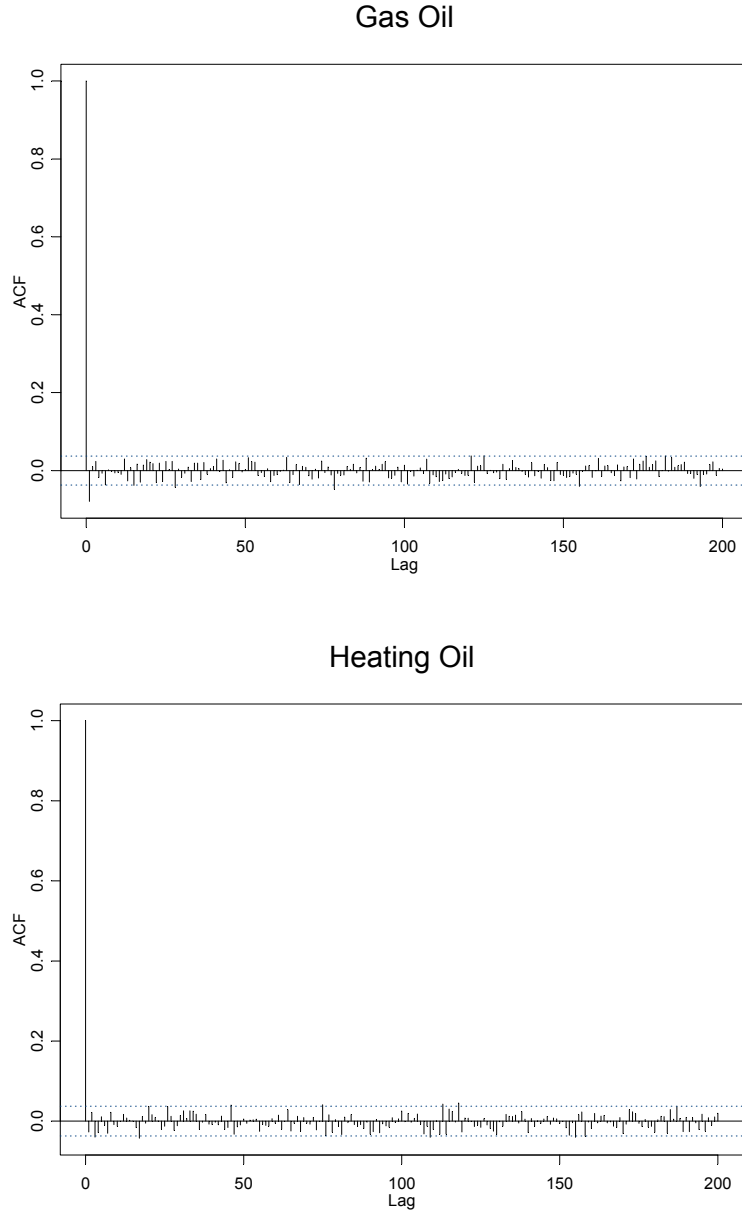


### 5.3.1 Time series dependence

We also look at time series dependence. Visual inspection of the sample ACF of  $\left(\hat{U}_{tk}\right)_{t \in \{1, \dots, n\}}$  and  $\left(\left|\hat{U}_{tk} - 1/2\right|\right)_{t \in \{1, \dots, n\}}$  does not seem to reveal any strong time dependence. For  $\left|\hat{U}_{tk} - 1/2\right|$ , in some occasions, (i.e. gas oil and oats) there is some mild first order dependence, but not necessarily positive. Figure IV plots the sample ACF of  $\left|\hat{U}_{tk} - 1/2\right|$  for gas oil and heating oil.

Figure IV.

Figure IV. Sample ACF for  $|U_t - 1/2|$



We also perform a test for the first order lag time series independence based on the  $L_p$  norm of  $\xi_n^{(2)}(\mathbf{u})$  (see (16)). The actual test statistic computed is given in (15). The critical values were obtained by 1000 simulations of (15) under the null using  $p = 2, \infty$ , for  $p$  as in (15). Table III reports the results.

Table III

Table III. Independence Test

	p=infinity	p=2
CRUDE OIL	0.91	0.07
GAS OIL (IPE)	0.90	0.08
HEATING OIL	1.47 *	0.15
NATURAL GAS	0.95	0.11
PROPANE	0.98	0.15
UNLEADED GAS	0.98	0.12
COCOA	0.90	0.09
COFFEE	0.84	0.07
SUGAR	0.83	0.06
ORANGE JUICE	0.62	0.05
SOYBEANS	1.19	0.13
CORN	0.70	0.08
OATS	0.56	0.03
WHEAT	0.78	0.08
COTTON	0.84	0.07

\* rejected at 90%

\*\* rejected at 95%

\*\*\* rejected at 99%

The null of serial independence does not seem to be rejected at one lag except for heating oil. This is surprising given our remarks about the visual inspection of the sample ACF's. A possible explanation for the large value in the test statistic of heating oil is that this commodity may fail to be uniform (see Table II). Recall that under  $H_0^{(1)}$  and  $H_1^{(1)}$  the random variables need to be uniform. Failing to be uniform invalidates the test, but does not affect inference on the sample ACF.

## 5.4 Cross-Sectional Analysis

As mentioned before, the data have been divided into three categories. For this reason, assuming a stationary covariance structure of the transformed uniform data, we report the sample correlation matrix of the uniform transforms within each group, but not among groups in Table IV.

Table IV.



Table IV. Correlation Matrices for Uniform Transforms

Energy Group						
	CRUDE OIL	GAS OIL (IPE)	HEATING OIL	NATURAL GAS	PROPANE	UNLEADED GAS
CRUDE OIL	1	0.46	0.79	0.20	0.44	0.74
GAS OIL (IPE)	-	1	0.53	0.13	0.43	0.42
HEATING OIL	-	-	1	0.25	0.48	0.71
NATURAL GAS	-	-	-	1	0.25	0.19
PROPANE	-	-	-	-	1	0.39
UNLEADED GAS	-	-	-	-	-	1

Miscellaneous Group				
	COCOA	COFFEE	SUGAR	ORANGE JUICE
COCOA	1	0.10	0.08	0.03
COFFEE	-	1	0.06	0.03
SUGAR	-	-	1	0.01
ORANGE JUICE	-	-	-	1

Grains/Oilseeds/Fiber Group					
	SOYBEANS	CORN	OATS	WHEAT	COTTON
SOYBEANS	1	0.57	0.43	0.38	0.13
CORN	-	1	0.52	0.52	0.10
OATS	-	-	1	0.39	0.07
WHEAT	-	-	-	1	0.09
COTTON	-	-	-	-	1

We then approximate the correlation matrix  $\hat{\Sigma}_t$  and apply the transform in (8) to derive a sequence of cross-sectionally independent random vectors. This implies that we orthogonalize the random variables in Gaussian space and transform them back into uniform. Table V reports the correlation matrix of the uniform random variables after this nonlinear transformation. The transformed series exhibits substantially less correlation within groups, though some correlation appears still to be present.

Table V.

Table V. Correlation Matrices of Uniform Transform

after the Transformation in (8)

Energy Group						
	CRUDE OIL	GAS OIL (IPE)	HEATING OIL	NATURAL GAS	PROPANE	UNLEADED GAS
CRUDE OIL	1	0.06	0.19	0.02	0.04	0.10
GAS OIL (IPE)	-	1	-0.02	0.00	0.05	-0.03
HEATING OIL	-	-	1	0.05	-0.01	0.07
NATURAL GAS	-	-	-	1	0.03	-0.01
PROPANE	-	-	-	-	1	-0.01
UNLEADED GAS	-	-	-	-	-	1

Miscellaneous Group				
	COCOA	COFFEE	SUGAR	ORANGE JUICE
COCOA	1	0.01	0.01	0.01
COFFEE	-	1	0.01	-0.01
SUGAR	-	-	1	0.00
ORANGE JUICE	-	-	-	1

Grains/Oilseeds/Fiber Group					
	SOYBEANS	CORN	OATS	WHEAT	COTTON
SOYBEANS	1	0.10	0.04	0.03	0.02
CORN	-	1	0.05	0.07	0.00
OATS	-	-	1	0.02	0.02
WHEAT	-	-	-	1	0.00
COTTON	-	-	-	-	1

In order to verify that the dependence structure defined by the Gaussian copula is satisfactory, we conducted the test described in Section 3.2 for  $p = 2, \infty$ . Results are reported in Table VI. The results confirm that meta-ellipticity with nonhomogeneous scaling matrix could be a suitable assumption to make (see Appendix B for few remarks on the simulated critical values and power).

Table VI

Table VI. Test for Meta-Elliptical Dependence

	p=infinity	p=2
Energy Group	0.90	0.02
Miscellaneous Group	1.23	0.04
Grains/Oilseeds/Fiber Group	1.09	0.03

\* rejected at 90%

\*\* rejected at 95%

\*\*\* rejected at 99%

## 5.5 Summary of Results

We briefly summarize some of the results.

- The tails of the log returns of commodity prices could be characterized by a time varying (symmetric) double Weibull distribution;
- There is a strong evidence of changes in the scale parameter of the distribution like in the case of financial returns;
- The shape parameter of the distribution is sub-Gaussian ( $\delta \simeq 2$  would imply nearly Gaussian tails), and perhaps larger than exponential;
- Cross-sectional dependence is quite strong in some cases, and could be time varying;
- Meta-ellipticity with a Gaussian generator might be a good first approximation everywhere apart from the extreme tails (see remarks about power in Appendix B). Hence, a meta-elliptical copula that exhibit tail dependence should be used (as advocated by many authors, e.g. Demarta and McNeil, 2005).

## 6 Final Remarks

We have considered forecast validation for time series that can be modelled by meta-elliptical distributions. In particular, we considered a framework where the distribution of the data conditional on past observations can be described by time varying parameters whose role is similar to the one of sufficient statistics. We gave several examples on the copula marginal decomposition. While not new, the copula decomposition requires some care when we model the conditional marginals and copula separately. In this case, the parameters of the copula cannot depend on past observations. This issue is often overlooked in many models like Engle's DCC model.

We proposed an inferential approach that followed the prequential principle in the sense that hypothesis were tested only on the basis of observables using recursive prediction. Since all parameters' values are calibrated using past observations only, under a properly specified null hypothesis, the forecast can be transformed into martingale differences which makes the performance of the model amenable to statistical analysis.

We considered the analysis of 16 energy and soft commodity returns as illustrative example. This example is also of interest in its own right because usually empirical analysis of returns has been confined to the case of financial indices. A simple model -chosen for illustrative purposes- could not be fully discredited by the data. There was evidence that the copula

specification could be discredited as it does not allow to capture tail dependence, which could be important when trying to forecast the joint distribution of commodity returns.

A natural application of these forecasting exercises is the calculation of value at risk, and related risk measures. The possibility of simulating from meta-elliptical distributions makes them viable and flexible candidates for real life applications. However, it is often argued that in financial applications, the dependence structure could be asymmetric (e.g. Kroner and Ng, 1998; see also Sancetta and Satchell, 2007, and references therein). This would require the extension to skewed elliptic distributions (e.g. Branco and Dey, 2001) and this is the subject of ongoing research.

## A Appendix: Online Forecast Combination

The combination weights used in the empirical section are obtained using variations of well known algorithms in the machine learning literature (Cesa-Bianchi and Lugosi, 2006). We only discuss the implementation, as it is beyond the scope of this paper to provide detailed explanations (the interested reader can consult the cited references).

To issue a marginal forecast, the weights are chosen to satisfy the following recursion: given  $w_{t-1,i}$ ,

$$\begin{aligned} w'_{ti} &: = w_{t-1,i} pdf_{|R_{t-1}||\mathcal{F}_{t-2}}(R_{t-1}|i) \\ w'_{ti} &: = w'_{ti} / \sum_{i=1}^4 w'_{ti} \\ w_{ti} &= (1 - \alpha) w_{ti} + \frac{\alpha}{3} \sum_{j \neq i} w_{tj} \end{aligned}$$

(see Cesa-Bianchi and Lugosi, 2006 and Sancetta, 2007, and references therein for the properties of this algorithm). We set  $\alpha = 10^{-3}$  and note that only observations up to time  $t - 1$  are used in the forecast and the initial weights are set equal to  $1/4$  (equal weighting). The combination weights for the covariance matrix are based on a similar recursion. Let  $(\hat{\mathbf{Z}}_t)_{t \in \mathbb{N}}$  be the sequence of Gaussian random vectors obtained from the uniforms  $(\hat{\mathbf{U}}_t)_{t \in \mathbb{N}}$  using  $\mathcal{Q}_\varphi$ . Let

$$g_{ti} := -2 \sum_{1 \leq k, l \leq K} \left[ \hat{\mathbf{Z}}_t \hat{\mathbf{Z}}_t^T - \sum_{i=1,2} \omega_{t-1,i} \hat{\Sigma}_t^{(i)} \right]_{kl} \hat{\Sigma}_{tkl}^{(i)},$$

which is the  $i^{th}$  entry in the gradient of the squared Frobenius norm of  $\left[ \hat{\mathbf{Z}}_t \hat{\mathbf{Z}}_t^T - \sum_{i=1,2} \omega_{ti} \hat{\Sigma}_t^{(i)} \right]$ ,

where  $k, l$  identifies its  $(k, l)$  entry. Then, given  $\omega_{t-1i}$ ,

$$\begin{aligned}
\omega'_{ti} &: = \omega_{t-1i} \exp \left\{ -t^{-1/2} g_{t-1i} \right\} \\
\omega'_{ti} &: = \omega'_{ti} / (\omega'_{t1} + \omega'_{t2}) \\
\psi &: = \sum_{i=1,2} I_{\{\omega'_{ti} < \gamma/2\}} \\
\omega'_{t\psi} &: = \sum_{i=1,2} \omega'_{ti} \\
\omega_{t+1,i} &= \omega'_{ti} [1 - \psi\gamma/2] / [1 - \omega'_{t\psi}] \text{ if } \omega'_{ti} \geq \gamma/2, \omega_{t+1,i} = \gamma/2 \text{ otherwise.}
\end{aligned}$$

We set  $\gamma = .01$ . Note that the update from  $\omega'_{ti}$  to  $\omega_{t+1,k}$  only makes sure that if  $\omega'_{ti} < \gamma/2$  then we set  $\omega_{t+1,k} = \gamma/2$ . After changing one weight, we need to adjust the remaining weight so that  $\omega_{t+1,1} + \omega_{t+1,2} = 1$  and the weights are positive. For the properties of this algorithm see Cesa-Bianchi and Lugosi (2006) and Sancetta (2006).

## B Appendix: Critical Values

In Table VII, we report critical values from 1000 simulations under the null of the true copula being the independence copula. The statistic computes (15) for  $p = 2$  and  $\infty$  with the obvious modifications.

Table VII

Confidence Level n=2570	Critical Value	
	p=infinity	p=2
<b>K=2</b>		
90%	1.440	0.255
95%	1.562	0.325
99%	1.825	0.468
<b>K=4</b>		
90%	1.444	0.084
95%	1.578	0.104
99%	1.773	0.138
<b>K=5</b>		
90%	1.351	0.044
95%	1.434	0.051
99%	1.638	0.070
<b>K=6</b>		
90%	1.184	0.022
95%	1.273	0.025
99%	1.507	0.034

To check power, we considered an alternative of a Gaussian copula with correlation matrix of .1 on the off diagonal entries (the null is a diagonal matrix). By Monte Carlo simulation, we computed the power when  $p = 2$  for the 95% critical values and this was found to be equal to about one except when  $K = 2$  (only dimensions  $K = 2, 4, 5, 6$  were considered). For  $K = 2$  the power is very low: about .28 . For  $p = \infty$  we have slightly less power and an extreme deterioration for  $K = 2$ : about .13. Hence caution has to be exercised carrying out inference for  $K = 2$ . We also considered an alternative of a t-copula with 6 degrees of freedom. The t-copula is considered as an important alternative to a Gaussian copula in modelling joint returns, as it exhibits tail dependence (e.g. Demarta and McNeil, 2005). A bivariate t-copula with 6 degrees of freedom has a coefficient of tail dependence equal to about .033: with probability of about 3%, we would observe a very low value of a variable given that the other variable takes a very low value. Under this alternative, the power is quite low: never above .8. The use of  $p = \infty$  produces a more powerful test. This is expected, as the difference between the alternative and the null mainly shows up in the tails. If we increased the degrees of freedom, the power would increase because the t-copula converges to the Gaussian copula. Hence, it might be convenient to standardize the test statistic by its theoretical variance and only consider a suitable subset of the hypercube bounded away from the boundary to avoid instability of the test statistic (recall the remark just before Section 3.1).

As a final remark, precise asymptotic critical values have been tabulated by Cotterill and Csorgo (1982) in this case. Differences between those critical values and the ones reported here are relatively small.

## References

- [1] Adler, R.J. (2000) On Excursion Sets, Tube Formulae and Maxima of Random Fields. *Annals of Applied Probability* 10, 1-74.
- [2] Adler, R and J.E. Taylor (2007) *Random Fields and Geometry*. New York: Springer, forthcoming.
- [3] Branco, M.D. and D.K. Dey (2001) A General Class of Skew-Elliptical Distributions. *Journal of Multivariate Analysis* 79, 99-113.

- [4] Breymann, W., A. Dias and P. Embrechts (2003) Dependence Structures for Multivariate High-Frequency Data in Finance. *Quantitative Finance* 3, 1-14.
- [5] Cesa-Bianchi, N. and G. Lugosi (2006) *Prediction, Learning , and Games*. Cambridge: Cambridge University Press.
- [6] Cheng, M.-Y., J. Fan and V. Spokoiny (2003) Dynamic Nonparametric Filtering with Application to Finance. Report 845, Weierstrass Institute for Applied Analysis and Stochastics.
- [7] Cotterill, D.S. and M. Csörgö (1982) On the Limiting Distribution of and Critical Values for the Multivariate Cramér-von Mises Statistic. *Annals of Statistics* 10, 233-244.
- [8] Davis, R.A. and T. Mikosch (1998) Limit Theory for the Sample ACF of Stationary Process with Heavy Tails with Application to ARCH. *Annals of Statistics* 26, 2049-2080.
- [9] Dawid, A.P. (1984) Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach. *Journal of the Royal Statistical Society Ser. A* 147, 278-292.
- [10] Dawid, A.P. (1985) Calibration-Based Empirical Probability. *The Annals of Statistics* 13, 1251-1274.
- [11] Dawid, A.P. (1986) Probability Forecasting. In S. Kotz, N.L. Johnson and C.B. Read (eds.), *Encyclopedia of Statistical Sciences* Vol. 7, 210-218. Wiley.
- [12] Dawid, A.P. (1997) Prequential Analysis. In S. Kotz, C.B. Read and D.L. Banks (eds.), *Encyclopedia of Statistical Sciences* Volume 1, 464-470. Wiley.
- [13] Dawid, A.P. and V. Vovk (1999) Prequential Probability: Principles and Properties. *Bernoulli* 5, 125-162.
- [14] Deaton A. and G. Laroque (1992) On the Behaviour of Commodity Prices. *The Review of Economic Studies* 59, 1-23.
- [15] Deaton A. and G. Laroque (1996) Competitive Storage and Commodity Price Dynamics. *The Journal of Political Economy* 104, 896-923.

- [16] Deb, P., P.K. Trivedi and P. Varangis (1996) The Excess Co-Movement of Commodity Prices Reconsidered. *Journal of Applied Econometrics* 11, 275-291.
- [17] Demarta, S. and A.J. McNeil (2005) The  $t$  Copula and Related Copulas. *International Statistical Review* 73, 111-129.
- [18] Diebold, F.X., T. Gunther and A. Tay (1998) Evaluating Density Forecasts, with Applications to Financial Risk Management. *International Economic Review* 39, 863-883.
- [19] Engle, R. (2002) Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business and Economic Statistics* 20, 339-350.
- [20] Fang, H.B., K.T. Fang and S. Kotz (2002) The Meta-Elliptical Distributions with Given Marginals. *Journal of Multivariate Analysis* 82, 1-16.
- [21] Fermanian, J.-D. (2005) Goodness of Fit Tests for Copulas. *Journal of Multivariate Analysis* 95, 119-152.
- [22] Fermanian, J.-D. and M. Wegkamp (2004) Time-Dependent Copulas. Preprint <[http://www.crest.fr/pageperso/fermanian/cond\\_copula10.pdf](http://www.crest.fr/pageperso/fermanian/cond_copula10.pdf)>
- [23] Frisch, U. and D. Sornette (1997) Extreme Deviations and Applications. *Journal de Physique I France* 7, 1155-1171.
- [24] Genest, C. and L.-P. Rivest (1993) Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association* 88, 1034-1043.
- [25] Genest C. and J.V. Zidek (1986) Combining Probability Distributions: A Critique and an Annotated Bibliography. *Statistical Science* 1, 114-148.
- [26] Ghosal, S., A. Sen and A.W. van der Vaart (2000) Testing Monotonicity of Regression. *Annals of Statistics* 28, 1054-1082.
- [27] Hansen, B.E. (1994) Autoregressive Conditional Density Estimation. *International Economic Review* 35, 705-730.
- [28] Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.



- [29] Hult, H. and F. Lindskog (2002) Multivariate Extremes, Aggregation and Dependence in Elliptical Distributions. *Advances in Applied Probability* 34, 587-608.
- [30] Joe, H. (1997) *Multivariate Models and Dependence Concepts*. London: Chapman & Hall Ltd.
- [31] Jondeau, E. and M. Rockinger (2005) The Copula-GARCH Model of Conditional Dependencies: An International Stock-Market Application. *Journal of International Money and Finance*, forthcoming.
- [32] Kano, Y. (1994) Consistency Property of Elliptical Probability Density Functions. *Journal of Multivariate Analysis* 51, 139-147.
- [33] Laherrère, J. and D. Sornette (1998) Stretched Exponential Distributions in Nature and Economy: "Fat Tails" with Characteristic Scales. *The European Physical Journal B* 2, 525-539.
- [34] Lindskog F., A. McNeil and U. Schmock (2003) Kendall's Tau for Elliptical Distributions. In G. Bol, G. Nakhaeizadeh, S.T. Rachev, T. Ridder and K.-H. Vollmer (editors), *Credit Risk: Measurement, Evaluation, and Management*, 149-156. Heidelberg: Springer.
- [35] Malevergne, Y. and D. Sornette (2003) Testing the Gaussian Copula Hypothesis for Financial Assets Dependences. *Quantitative Finance* 3, 231-250.
- [36] Mercurio, D. and V. Spokoiny (2004) Statistical Inference for Time-Inhomogeneous Volatility Models. *The Annals of Statistics* 32, 577-602.
- [37] Mikhaleva, T.L. and V.I. Piterbarg (1996). On the Distribution of the Maximum of a Gaussian Field with Constant Variance on a Smooth Manifold. *Theory of Probability and its Applications* 41, 367-379.
- [38] Muth J.F. (1960) Optimal Properties of Exponentially Weighted Forecasts *Journal of the American Statistical Association* 290, 299-306.
- [39] Patton, A. (2006) Modelling Asymmetric Exchange Rate Dependence. *International Economic Review* 47, 527-556.

- [40] Pesaran, M.H. and A. Timmermann (2005) Real Time Econometrics. *Econometric Theory* 21, 212-231.
- [41] Pindyck R.S. (2004) Volatility and Commodity Price Dynamics. *Journal of Futures Markets* 24, 1029-1047.
- [42] Rio, E. (2000) *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*. Paris: Springer.
- [43] Romano, J.P., A.M. Shaikh and M. Wolf (2008) Formalized Data Snooping Based on Generalized Error Rates. Forthcoming in *Econometric Theory*.
- [44] Rüschendorf, L. and V. de Valk (1993) On Regression Representation of Stochastic Processes. *Stochastic Processes and their Applications* 46, 183-198.
- [45] Sancetta, A. (2006) Online Forecast Combination for Dependent Heterogeneous Data. Preprint.
- [46] Sancetta, A. (2007) Online Forecast Combinations of Distributions: Worst Case Bounds. *Journal of Econometrics* 141, 621-651.
- [47] Sancetta, A. and S.E. Satchell (2007) Changing Correlation and Equity Portfolio Diversification Failure for Linear Factor Models During Market Declines. *Applied Mathematical Finance* 14, 227-242.
- [48] Seillier-Moiseiwitsch, F. and A.P. Dawid (1993) On Testing the Validity of Sequential Probability Forecasts. *Journal of American Statistical Association* 88, 355-359.
- [49] Silvapulle P. and C.W.J. Granger (2001) Large Returns, Conditional Correlation and Portfolio Diversification: A Value-at-Risk Approach. *Quantitative Finance* 10, 542-551.
- [50] Sklar, A. (1973) Random Variables, Joint Distribution Functions, and Copulas. *Kybernetika* 9, 449-460.
- [51] Stărică, C. and C.W.J. Granger (2005) Non-Stationarities in Stock Returns. *Review of Economics and Statistics*, forthcoming.
- [52] Taylor, S.J. (1980) Conjectured Models for Trends in Financial Prices, Tests and Forecasts. *Journal of the Royal Statistical Society, Series A*, 143, 338-362.

- [53] Timmermann, A. (2006) Forecast Combinations. Forthcoming in G. Elliott, C.W.J Granger and A. Timmermann (eds.) Handbook of Economic Forecasting. North Holland.
- [54] Van Garderen K.J. (1997) Curved Exponential Models in Econometrics. *Econometric Theory* 13, 771-790.
- [55] Van der Vaart, A. and J.A. Wellner (2000) Weak Convergence of Empirical Processes. Springer Series in Statistics. New York: Springer.
- [56] Yang, Y. (2004) Combining Forecasting Procedures: Some Theoretical Results. *Econometric Theory* 20, 176-222.